JOURNAL OF MULTIVARIATE ANALYSIS 11, 354–367 (1981)

Large Deviation Probabilities for Certain Dependent Processes*

KESAR SINGH†

Indian Statistical Institute, Calcutta, and Stanford University

Communicated by P. R. Krishnaiah

Certain results on large deviation probabilities for linear and m-dependent processes are considered here.

1. INTRODUCTION

This work is a contribution towards relaxing independence in the theory of LDP. Section 2 of this paper contains a Chernoff's theorem type result for linear processes under the absolute summability condition of the coefficients. Sections 3 and 4 prove a kind of limit theorem for various statistics based on m-dependent processes. The relevance of the later limit theorems is explained below.

Let $\{\hat{\mu}_n\}$ be a stochastic sequence intending to estimate the real parameter μ. Consider the root of the equation

$$2(1-\Phi(\varepsilon/\gamma_{n,\varepsilon}))=P(|\mu_n-\mu|\geqslant\varepsilon),$$

where ε is some positive number. If there is a sequence $\gamma(n)$ of positive numbers s.t.

$$\liminf_{\varepsilon\to0}\liminf_{n\to\infty}\gamma^2_{n,\varepsilon}/\gamma(n)=1=\limsup_{\varepsilon\to0}\limsup_{n\to\infty}\gamma^2_{n,\varepsilon}/\gamma(n)$$

then $\gamma(n)$ is called AEV (asymptotic effective variance) of μ_n (see Bahadur, 1960). If for all positive sequences $\varepsilon_n\to\varepsilon>0$,

$$n^{-1}\log P(|\mu_n-\mu|\geqslant\varepsilon_n)=-(\varepsilon^2/2\gamma)(1+O_\varepsilon+O_{n,\varepsilon}),\qquad(1.1)$$

Received: May 18, 1979; revised: April 3, 1981.
AMS 1980 subject classification: Primary 60F10.
Key words and phrases: Large deviations, dependent processes.
* This work was partially supported under a Ford Foundation grant.

† Currently at Rutgers University.

0047-259X/81/030354–14$02.00/0

where $\gamma > 0$, $\lim_{\epsilon\to 0} O_\epsilon = 0$ and $\lim_{n\to\infty} O_{n,\epsilon} = 0$ for all ε sufficiently small, then γ/n is an AEV of μ_n and it is unique in the sense that for any other sequence $\gamma'(n)$ with the same property, $\lim_{n\to\infty} \gamma/n\gamma'(n) = 1$.

An estimate like (1.1) can be interpreted in testing of hypothesis problems as follows. Let the testing problem be $\theta = \theta_0$ vs $\theta > \theta_0$, where θ is some real parameter of a statistical family. Let the test statistic be $T_n - D(\theta_0)$, large values of the statistic being significant. The level attained is $1 - Q_n(T_n - D(\theta_0))$, where Q_n is the d.f. of $T_n - D(\theta_0)$ under θ_0.

Assume that, under θ_0,

$$n^{-1} \log P(T_n - D(\theta_0) \geqslant \varepsilon_n) = -(\varepsilon^2/2\gamma_0)(1 + O_\epsilon + O_{n,\epsilon}),$$

where $\gamma_0 > 0$, $\varepsilon_n \to \varepsilon$, and O_ϵ and $O_{n,\epsilon}$ are as in (1.1). Further, let $T_n \to D(\theta)$ a.s. $[\theta]$ with $D(\theta)$ s.t. $D(\theta) > D(\theta_0)$ for $\theta > \theta_0$ and $D(\theta) \to D(\theta_0)$ as $\theta \searrow \theta_0$. Then, under $\theta > \theta_0$,

$$n^{-1} \log(1 - Q_n(T_n - D(\theta_0)) = -[D(\theta) - D(\theta_0)]^2(1 + r_\theta + r_{n,\theta})/2\gamma_0,$$

where $\lim_{\theta\searrow\theta_0} r_\theta = 0$ and $\lim_{n\to\infty} r_{n,\theta} = 0$ a.s. $[\theta]$ for all θ sufficiently close to θ_0. In a sense, this means that the performance of a test based on $T_n - D(\theta_0)$ is proportional to $(D(\theta) - D(\theta_0))^2$ and inversely proportional to γ_0, locally. A similar interpretation holds for two-sided tests as well. In Sections 3 and 4 we establish bounds like (1.1) for various statistics based on m-dependent processes.

2. Sample Mean of Linear Processes

We derive a Chernoff's theorem type LDP result for linear processes using the following general theorem. For a r.v. X, let $\sup(X) = \inf\{x: P(X > x) = 0\}$.

THEOREM 1. *Let $\{T_n\}$ be a sequence of statistics having the representation $|T_n - \sum_{i=1,n} Y_i| \leqslant R_n$, where $\{Y_i\}$ is a sequence of i.i.d. r.v.'s having mean zero and R_n is another sequence of statistics. For $\alpha > 0$, let*

$$h(n, \alpha) = \inf\{x \geqslant 0 \text{ s.t. } E\exp(xR_n) \geqslant \exp(\alpha n)\}$$
$$= n \qquad \text{if } R_n \text{ is degenerate at zero.}$$

If for some α finite, $\lim_{n\to\infty} h(n, \alpha) = \infty$, *then*

$$\lim_{n\to\infty} n^{-1} \log P(T_n \geqslant nd) = \inf_{t\geqslant 0}\{-td + \log E\exp(tY_1)\}$$

$$(=0 \text{ if } Y_1 \text{ does not have finite m.g.f.}) \qquad (2.1)$$

for $d > 0$, *provided one of the following holds*: A_1. $d \neq \sup(Y_1)$. A_2. $P(Y_1 = \sup(Y_1)) = 0$.

Proof. It follows from the proof of Lemma 3.3 of Bahadur (1969) that the function $g_x = \inf_{t>0} E \exp(-tx + tY_1)$ is continuous on $(0, \sup(Y_1))$. On $(\sup(Y_1), \infty)$, $g_x = 0$. Further, if A_2 holds, it can be shown that g_x is continuous on $(0, \infty)$. If g_x is continuous at d and $\lim_{n\to\infty} h(n, \alpha) = \infty$, it follows that

$$\lim_{n\to\infty} n^{-1} \log P(\bar{Y}_n \geqslant d \pm 4\alpha/h(n, \alpha)) = \log g_d. \tag{2.2}$$

Furthermore, some elementary inequalities and the definition of $h(n, \alpha)$ imply

$$\begin{aligned} &|P(T_n \geqslant nd) - P(\bar{Y}_n \geqslant d \pm 4\alpha/h(n, \alpha))| \\ &\quad \leqslant P(R_n \geqslant 4n\alpha/h(n, \alpha)) \\ &\quad \leqslant \exp(-2\alpha n) E \exp(h(n, \alpha) R_n/2) \leqslant \exp(-\alpha n). \end{aligned} \tag{2.3}$$

Since the function $h(n, \alpha)$ is nondecreasing in α for every fixed n, α can be chosen to be arbitrarily large; hence (2.2) and (2.3) yield (2.1).

We state the LDP result for linear processes as

COROLLARY 2.1. *Let* $\{Z_i\}$ *be a double sequence of i.i.d. r.v.'s with mean zero and* $E \exp(tZ_0) < \infty$ *for all real* t. *Let the process* $\{X_i\}$ *be defined as* $X_i = \sum_{j=0,\infty} a_j Z_{i-j}$, *where* $\sum_{j=0,\infty} |a_j| < \infty$ *and* $z = \sum_{j=0,\infty} a_j \neq 0$. *Then*

$$\lim_{n\to\infty} n^{-1} \log P(\bar{X}_n \geqslant d) = \inf_{t \geqslant 0} \{-td + \log(E \exp(tzZ_0))\} \tag{2.4}$$

for $d > 0$ *provided one of the following holds*: (A_1^*) $d \neq \sup(zZ_0)$, (A_2^*) $P(zZ_0 = \sup(zZ_0)) = 0$.

Proof. Let us define $X_{ii} = \sum_{j=0,i-1} a_j Z_{i-j}$ and $t_{a,b} = \sum_{i=a,b} a_i$ where a, b are nonnegative integers and $b(\geqslant a)$ can be $+\infty$. Let

$$z_0 = 1 + \sup_{i \geqslant 1} \sum_{j=i,\infty} |a_j|,$$

so that $z_0 > 0$ and $|t_{a,b}|/z_0 \leqslant 1$ for all $1 \leqslant a \leqslant b \leqslant \infty$. Evidently $\sum_{i=1,n} X_{ii} = \sum_{i=1,n} t_{0,n-i} Z_i = \sum_{i=1,n} (zZ_i - t_{n-i+1,\infty} Z_i)$.

Therefore it follows that

$$n |\bar{X}_n - z\bar{Z}_n| \leqslant R_n = \sum_{i=1,n} |t_{n-i+1,\infty} Z_i| + \sum_{i=0,\infty} |t_{i+1,n+i} Z_{-i}|.$$

Applying the fact that Lp-norms are nondecreasing in $p \geqslant 1$,

$$\log(E\exp(tR_n)) = \log(E\exp(tz_0(R_n/z_0)))$$
$$\leqslant \left[\sum_{i=1,n} |t_{n-i+1,\infty}|/z_0 + \sum_{i=0,\infty} |t_{i+1,n+i}|/z_0\right] \log(E\exp(tz_0\,|Z_0|)). \quad (2.5)$$

If $\sum_{i=0,\infty} |a_i| < \infty$, both the terms in the [ ] bracket above are $o(n)$, so that

$$E\exp(tR_n) \leqslant (E(\exp(z_0 t\,|Z_0|)))^{b_n},$$

where $n/b_n \to \infty$ as $n \to \infty$. This implies in this case that

$$h(n, 1) \geqslant \inf\{x \text{ s.t. } E\exp(z_0 x\,|Z_0|) \geqslant \exp(n/b_n)\}$$

which converges to ∞ as $n \to \infty$; thus this corollary follows from Theorem 1. ∎

It may not be out of place to mention that Chanda (1972) also established a LDP result for linear processes assuming that $a_i \ll \rho^i$ for some $\rho \in (0, 1)$ and that the distribution of Z_0 satisfies Cramér's condition. We give below a counterexample to show that if none of the conditions (A_1^*) and (A_2^*) hold then (2.4) may be false.

COUNTEREXAMPLE. Let $0 < d = \sup(Z_0) < \infty$, $0 < P(Z_0 = d) < 1$, and $a_i = 2^{-i-1}$ for $i \geqslant 0$. Then $z = \sum_{i=0,\infty} a_i = 1$ and

$$P(\bar{X}_n \geqslant d) = P(Z_i = d \text{ for all integers } i \leqslant n) = 0,$$

whereas

$$P(\bar{Z}_n \geqslant d) = P(Z_i = d \text{ for all } 1 \leqslant i \leqslant n) = (P(Z_0 = d))^n;$$

so that $\inf_{t>0}\{-td + \log E\exp(tZ_0)\} = \log P(Z_0 = \sup(Z_0)) > -\infty$.

3. LINEAR STATISTICS BASED ON m-DEPENDENT PROCESSES

Hereafter $\{Y_i\}$ denotes a sequence of m-dependent r.v.'s with mean 0, c_i's are positive constants. ε_n denotes a sequence with the property that $\lim_{n\to\infty} \varepsilon_n = \varepsilon$. O_ϵ and $O_{n,\epsilon}$ are remainders with the properties $\lim_{\epsilon\to 0} O_\epsilon = 0$ and $\lim_{n\to\infty} O_{n,\epsilon} = 0$ for all ε sufficiently small. For convenience, let us say that a process $R(n, \varepsilon)$ is (*) if for some δ_1, δ_2 positive

$$P(|R(n, \varepsilon)| \geqslant \varepsilon^{1+\delta_1}) \leqslant c_1 \exp(-\varepsilon^{2-\delta_2} n(1 + O_\epsilon + O_{n,\epsilon}))$$

for all positive ε, sufficiently small. If a statistic T_n has the decomposition $T_n = T_{n,\epsilon} + R_{n,\epsilon}$, $R_{n,\epsilon}$ is (*), and $T_{n,\epsilon}$ has the LDP bound

$$n^{-1} \log P(T_{n,\epsilon} \geqslant \varepsilon_n) = -(\varepsilon^2/2\gamma)(1 + O_\epsilon + O_{n,\epsilon}),$$

then so does T_n.

THEOREM 2. *Let* $v = V(Y_1) + 2\sum_{i=1,m-1} \operatorname{cov}(Y_1, Y_{1+i}) > 0$ *and* $H(t) = E\exp(tY_i) < \infty$ *for all* $|t| \leqslant a > 0$. *Then*

$$n^{-1} \log P(\bar{Y}_n \geqslant \varepsilon_n) = -(\varepsilon^2/2v)(1 + O_\epsilon + O_{n,\epsilon}).$$

Proof. W.l.a.g. we assume $v = 1$. Let ε be small enough s.t. $\varepsilon^{-1/2} \geqslant m$. Taking $p = [\varepsilon^{-1/2}]$ and $k = [n/(p+m)]$, we break the sum $\sum_{i=1,n} Y_i$ as follows

$$\sum_{i=1,n} Y_i = \sum_{i=1,k} (\xi_i + \eta_i) + \xi_{k+1},$$

where $\xi_i = \sum_{j=1,p} Y_{(p+m)(i-1)+j}$ and $\eta_i = \sum_{j=1,m} Y_{(p+m)(i-1)+p+j}$. We shall see first that $n^{-1}\sum_{i=1,k}\eta_i$ and $n^{-1}\xi_{k+1}$ are (*). It follows by repeated application of Hölder's inequality that $E\exp(a|\eta_1|/m) \leqslant E\exp(a|Y_1|) < \infty$; therefore $E\exp(\varepsilon\eta_1) \leqslant 1 + c_2\varepsilon^2$ for all positive ε in a neighborhood of zero. Consequently

$$\begin{aligned} P\left(\sum_{i=1,k} \eta_i \geqslant 2c_2 n\varepsilon^{9/8}\right) &\leqslant \exp(-2c_2 n\varepsilon^{7/4})(E\exp(\varepsilon^{5/8}\eta_1))^k \\ &\leqslant \exp(-2c_2 n\varepsilon^{7/4})(1 + c_2\varepsilon^{5/4})^k \\ &\leqslant \exp(-2c_2 n\varepsilon^{7/4} + c_2 k\varepsilon^{5/4}) \\ &\leqslant \exp(-c_2 n\varepsilon^{7/4}) \end{aligned}$$

for all ε small. A similar bound is established for $-\sum_{i=1,k}\eta_i$ to conclude that $n^{-1}\sum_{i=1,k}\eta_i$ is (*). Turning to ξ_{k+1}, we note that

$$\begin{aligned} P(|\xi_{k+1}| \geqslant n\varepsilon^{9/8}) &\leqslant P\left(\sum_{i=1,p+m} |Y_i| \geqslant n\varepsilon^{9/8}\right) \\ &\leqslant \exp(-\varepsilon^{15/8} n) E\exp\left(\varepsilon^{3/4} \sum_{i=1,p+m} |Y_i|\right) \\ &\leqslant \exp(-\varepsilon^{15/8} n) E\exp(\varepsilon^{3/4}([\varepsilon^{-1/2}] + m)|Y_1|) \\ &= \exp(-\varepsilon^{15/8} n(1 + o(1))) \end{aligned}$$

if ε is small enough s.t. $\varepsilon^{3/4}([\varepsilon^{-1/2}]+m)\leqslant a$. Thus it only remains to be seen that

$$P\left(\sum_{i=1,k}\xi_i\geqslant n\varepsilon_n\right)=\exp(-n\varepsilon^2(1+O_\varepsilon+O_{n,\varepsilon})/2).$$

Let us write $P(\sum_{i=1,k}\xi_i\geqslant n\varepsilon_n)=P(\sum_{i=1,k}\xi'_i\geqslant ka_n)$, where $\xi'_i=\xi_i\varepsilon^{1/4}$, $a_n=(n/k)\,\varepsilon^{1/4}$, and $\varepsilon_n=\varepsilon_n\varepsilon^{-1/4}(1+O(\varepsilon^{1/2}))$. Let us define $H^*(s)=E\exp(s\xi'_1)$ and

$$\nu(s)=(H^*(s))^{-1}\int_{-\infty}^{\infty}x\exp(sx)\,dP(\xi'_1\leqslant x).$$

$H^*(s)$ and $\nu(s)$ are well defined in the region $|s|\leqslant a\varepsilon^{1/4}/4$, to which we confine the rest of the estimation. Applying Hölder's inequality and the moment inequality given as Lemma 1.9 of Ibragimov (1962), we have

$$\begin{aligned}&|H^*(s)-1-s^2E(\xi'_1)^2/2|\\&\quad\leqslant s^3E(|\xi'_1|^3\exp(s\,|\xi'_1|))/6\\&\quad\leqslant s^3\left[E\,|\xi'_1|^6E\exp\left(2s\varepsilon^{1/4}\sum_{i=1,p}|Y_i|\right)\right]^{1/2}\Big/6\\&\quad\leqslant s^3[E\,|\xi'_1|^6E\exp(2s\varepsilon^{1/4}p\,|Y_1|)]^{1/2}/6\leqslant c_3s^3\end{aligned}$$

in the region $|s|\leqslant a\varepsilon^{1/4}/4$. Also, it is easily seen that $E(\xi'_1)^2=1+O(\varepsilon^{1/2})$. Thus, $|H^*(s)-1-s^2/2|\leqslant c_4s^2\varepsilon^{1/4}$. By similar expansions, one finds that

$$|\nu(s)-s|\leqslant c_5s\varepsilon^{1/4}. \tag{3.1}$$

That the function $\nu(s)$ is continuous on $[0,a\varepsilon^{1/4}/4]$ and is right continuous at zero ($\nu(0)=0$) follows from the dominated convergence theorem. Utilizing these facts, (3.1), and the intermediate value theorem, we conclude that the equation $\nu(b_n)=a_n$ has solutions for all n large enough and that any solution has to satisfy $b_n=a_n+O(a_n\varepsilon^{1/4})$.

Now, define the d.f. $G(\cdot,b_n)$ by

$$dG(x,b_n)=(H^*(b_n))^{-1}\exp(b_nx)\,dP(\xi'_1\leqslant x)$$

and let $\sigma^2(b_n)$ and $r(b_n)$ be its variance and the third moment. If $G_n(\cdot,b_n)$ denotes the convolution of n copies of $G(\cdot,b_n)$, then

$$\begin{aligned}&P\left(\sum_{i=1,k}\xi'_i\geqslant ka_n\right)\\&\quad=\int_{ka_n}^{\infty}(H^*(b_n))^k\exp(-b_nx)\,dG_n(x,b_n)\\&\quad=\int_0^{\infty}A(b_n)\exp(-D(b_n)z)\,dG_n(\sigma(b_n)\,k^{1/2}z+k\nu(b_n),b_n),\end{aligned} \tag{3.2}$$

where $A(b_n) = (H^*(b_n))^k \exp(-b_n k a_n)$ and $D(b_n) = \sigma(b_n)\, b_n k^{1/2}$. Some easy expansions also show that $\sigma^2(b_n) = 1 + O(b_n)$, $A(b_n) = \exp(-k(b_n^2/2 + O(b_n^{5/2})))$, $D(b_n) = k^{1/2}(b_n + O(b_n^2))$, and $r(b_n) = O(1)$. The Berry–Esséen theorem and these estimates imply

$$\sup_{z \in \mathbb{R}} |G_n(\sigma(b_n)\, k^{1/2} z + k\nu(b_n), b_n) - \Phi(z)| \leqslant c_6 k^{-1/2}.$$

Now we are ready to estimate the desired probability. The r.h.s. of (3.2) equals

$$\begin{aligned} &A(b_n) \int_0^\infty \exp(-D(b_n) z)\, d\Phi(z) \\ &\qquad + A(b_n) \int_0^\infty \exp(-D(b_n) z)\, d(G_n(\sigma(b_n)\, k^{1/2} z) + k a_n, b_n) - \Phi(z) \\ &\quad = A(b_n) \exp((D(b_n))^2/2)[1 - \Phi(D(b_n))] + A(b_n)\, O(k^{-1/2}) \end{aligned}$$

(using the integration by parts)

$$= \exp(-n\varepsilon^2(1 + O_\epsilon + O_{n,\epsilon})/2),$$

substituting the estimates of $A(b_n)$, $D(b_n)$, and b_n.

Theorem 2 generalizes very easily to the case of linear processes. In fact, this extension only needs the assumption that the original r.v.'s have m.g.f. in a neighborhood of zero. Using the bound (2.5) it is seen that R_n is (*).

We now state a uniform version of Theorem 2. We need to recall this result in the next section while dealing with multivariate Kolmogorov–Smirnov statistics.

THEOREM 3. *Let f_μ, $\mu \in \Lambda$, be a family of measurable functions defined on real line. Let us assume* $\sup_{\mu \in \Lambda} |f_\mu(Y_i)| \leqslant f(Y_i)$, *where f is a measurable function s.t. $f(Y_1)$ has m.g.f. in a neighborhood of zero, and* $\inf_{\mu \in \Lambda} V_\mu > 0$, *where*

$$V_\mu = V(f_\mu(Y_1)) + 2 \sum_{i=1,m-1} \operatorname{cov}(f_\mu(Y_1), f_\mu(Y_{1+i})).$$

Then

$$\begin{aligned} &\sup_{\mu \in \Lambda} \left| n^{-1} \log P\left(n^{-1} \sum_{i=1,n} f_\mu(Y_i) - Ef_\mu(Y_i) \geqslant \varepsilon_n\right) + (\varepsilon^2/2V_\mu) \right| \\ &\quad = O_\epsilon + O_{n,\epsilon}. \end{aligned}$$

This claim is verified by checking that all the estimates in the proof of Theorem 2, when adopted for $f_\mu(Y_i)$, are uniform in $\mu \in \Lambda$ under the conditions assumed above.

The next theorem is a multivariate extension of Theorem 2. A corollary of this theorem has been applied in the next section in connection with the Anderson–Darling statistic.

THEOREM 4. *Let* $\{\mathbf{Y}_i\} = \{Y_{i1},\dots, Y_{iq}\}$ *be a stationary sequence of q-variate, m-dependent random vectors. We assume that* $\Sigma = \{\sigma_{ij}\}_{q\times q}$, *where*

$$\sigma_{ij} = \operatorname{cov}(Y_{1i}, Y_{1j}) + \sum_{k=1,m-1} (\operatorname{cov}(Y_{1i}, Y_{1+k,j}) + \operatorname{cov}(Y_{(1+k)i}, Y_{1j})),$$

is positive definite. For a bounded convex set $C \in \mathbb{R}^q$, *with the null vector as an interior point, let* $I_C = \inf\{h^2(\mathbf{a})/\mathbf{a}\Sigma\mathbf{a}'$, *where* $\mathbf{a}$ *is any nonnull vector and* $\mathbf{a}\cdot\mathbf{x} = h(\mathbf{a})$ *is a supporting hyperplane of* $C\}$. *Then*

$$n^{-1}\log P\left(n^{-1}\sum_{i=1,n}\mathbf{Y}_i - E(\mathbf{Y}_1) \in \varepsilon_n C^c\right) = -(\varepsilon^2 I_C)(1 + O_\epsilon + O_{n,\epsilon})/2,$$

where C^c *is the complement of the set* C.

The proof we have in mind is based on the ideas of Theorem 9 of Rubin and Sethuraman (1965). It turns out that, given our Theorems 2 and 3, essentially the same steps go through.

COROLLARY 3.1. *In the special case* $C = \{\mathbf{x}: \mathbf{x}A\mathbf{x}' \leqslant 1\}$, *where A is some positive definite matrix,* $(I_C)^{-1} = (A\Sigma)_* =$ *the largest eigenvalue of the matrix* $A\Sigma$. *Thus, if* $E\mathbf{Y}_1 = 0$,

$$n^{-1}\log P(\bar{\mathbf{Y}}_n A\bar{\mathbf{Y}}_n' \geqslant \varepsilon_n^2) = -(n\varepsilon^2/2(A\Sigma)_*)(1 + O_\epsilon + O_{n,\epsilon}),$$

where $\bar{\mathbf{Y}}_n = n^{-1}\sum_{i=1,n}\mathbf{Y}_i$. *Specializing further,*

$$n^{-1}\log P(\|\bar{\mathbf{Y}}_n\| \geqslant \varepsilon_n) = -(n\varepsilon^2/2(\Sigma)_*)(1 + O_\epsilon + O_{n,\epsilon}).$$

The last theorem of this section is about trimmed L-statistics. The idea of Bahadur–Kiefer representation of quantiles is exploited in the proof. Let $F_n(\cdot)$ and $F_n^{-1}(\cdot)$ denote the right continuous versions of empirical and quantile processes of $(Y_1, Y_2,\dots, Y_n)$.

THEOREM 5. *Let F be the d.f. of* Y_1 *and W be some function of bounded variation on* $[\alpha, \beta]$, $0 < \alpha \leqslant \beta < 1$. *We define*

$$L_n = \int_\alpha^\beta F_n^{-1}(t)\, dW(t), \qquad L = \int_\alpha^\beta F^{-1}(t)\, dW(t),$$

and assume that $F''(x)$ *exists* $\forall x \in (F^{-1}(\alpha - \delta), F^{-1}(\beta + \delta))$, *for a* $\delta > 0$; $0 < c_7 \geqslant F'(x) \geqslant c_8 > 0$ *and* $|F''(x)| \leqslant c_9$ *throughout the interval. If* $F'(F^{-1}(t)) = f_t$, $u_{it} = \{t - I(X_i \leqslant F^{-1}(t))\}/f_t$, $u_i = \int_\alpha^\beta u_{it}\, dW(t)$ *and*

$$0 < \sigma_L^2 = V(u_1) + 2 \sum_{i=1,m-1} \operatorname{cov}(u_1, u_{1+i}),$$

then

$$n^{-1} \log P(L_n - L > \varepsilon_n) = -(n\varepsilon^2/\sigma_L^2)(1 + O_\epsilon + O_{n,\epsilon})/2.$$

The proofs of this theorem and the theorems in the next section use an exponential bound which we present as

LEMMA 3.1. *Let* $\{X_i\}$ *be a stationary sequence of m-dependent, bounded r.v.'s with mean zero and* $V(X_1) + 2\sum_{i=1,m-1} \operatorname{cov}(Y_1, Y_{1+i}) \leqslant B \leqslant 1$. *For positive numbers* D, Z *satisfying* $Z \leqslant D$ *and* $ZnB \leqslant D^2$, *there exist positive constants* c_{10} *and* c_{11}, *depending upon m and* $\|X_1\|_\infty$ *only, s.t.*

$$P\left(\left|\sum_{i=1,n} X_i\right| \geqslant 2c_{10}D\right) \leqslant c_{11}e^{-Z}.$$

Proof. We write $\sum_{i=1,n} X_i = \sum_{i=1,k}(\xi_i + \eta_i) + \xi_{k+1}$ just as in the proof of Theorem 2, replacing Y_i by X_i and taking $p = m$.

$$\begin{aligned} P\left(\sum_{i=1,k+1} \xi_i \geqslant c_{10}D\right) &\leqslant \exp(-c_{10}Z) E \exp\left(ZD^{-1} \sum_{i=1,k+1} \xi_i\right) \\ &\leqslant c_{11} \exp(-c_{10}Z)(E\exp(ZD^{-1}\xi_1))^k \\ &\leqslant c_{11} \exp(-c_{10}Z + k\log(1 + c_{12}Z^2D^{-2}B)) \\ &\leqslant c_{11} \exp(-Z) \end{aligned}$$

for c_{10} large enough, since $ZnB \leqslant D^2$ and $V(\xi_1) \leqslant 2Bm$. $-\sum_{i=1,k+1} \xi_i$ and $\pm\sum_{i=1,k} \eta_i$ are treated similarly to get the bound.

Proof of Theorem 5. With Bahadur–Kiefer representation of quantiles in mind, we write

$$L_n - L = n^{-1} \sum_{i=1,n} u_i + R(n).$$

Under the assumed regularity conditions about F, if $E_n(\cdot)$ and $E_n^{-1}(\cdot)$ denote the right continuous versions of empirical and quantile processes of $U_i = F(X_i)$, $i = 1, \ldots, n$, then

$$
\begin{aligned}
|R(n)| &\leqslant c_{12} \sup_{t\in[\alpha,\beta]} |F_n^{-1}(t) - F^{-1}(t) + (F_n(F^{-1}(t)) - t) f_t^{-1}| \\
&= c_{12} \sup_{t\in[\alpha,\beta]} |F^{-1}(E_n^{-1}(t)) - F^{-1}(t) + (F_n(F^{-1}(t)) - t) f_t^{-1}| \\
&\leqslant c_{13} \sup_{t\in[\alpha,\beta]} |E_n^{-1}(t) - t + E_n(t) - t| \\
&\quad + c_{14}(\sup_{t\in[\alpha,\beta]} |E_n^{-1}(t) - t|)^2.
\end{aligned}
$$

The theorem is established by showing that the two statistics in the r.h.s. above are (*). This is done by the method of subdivision, applying Lemma 3.1 for probability bounds. The details are long but seem to be quite standard ones and so are omitted. Babu and Singh (1978) may be found helpful in verifying the claims.

4. Goodness of Fit Statistics

Throughout this section $\{\mathbf{Y}_i\}$ denotes a stationary sequence of q-variate, m-dependent random vectors with the underlying d.f. $\mathbf{F}$ having continuous marginals. We discuss in this section LDP for multivariate versions of two goodness of fit statistics—Kolmogorov–Smirnov statistics and Anderson–Darling statistics, defined as

$$KS_n = \sup\{|\mathbf{F}_n(\mathbf{x}) - \mathbf{F}(\mathbf{x})|; x \in \mathbb{R}^q\}$$

$$AD_n^2 = \int_{\mathbb{R}^q} [\mathbf{F}_n(\mathbf{x}) - \mathbf{F}(\mathbf{x})]^2 \, h^2(\mathbf{x}) \, d\mathbf{x},$$

where $\mathbf{F}_n$ denotes e.d.f. $\{\mathbf{Y}_i\}$ and h is some bounded continuous function on $\mathbb{R}^q$. We assume further that marginal distributions are $U[0, 1]$, which of course can be achieved by suitable transformation. This assumption reduces all our considerations to $[0, 1]^q$. We denote this set by J hereafter.

THEOREM 6. *If $\Gamma(\cdot, \cdot)$ denotes the covariance kernel on $J \times J$ defined as*

$$
\begin{aligned}
\Gamma(\mathbf{s}, \mathbf{t}) = {} & \operatorname{cov}(I(\mathbf{Y}_1 \leqslant \mathbf{s}), I(\mathbf{Y}_1 \leqslant \mathbf{t})) \\
& + \sum_{i=1,m-1} [\operatorname{cov}(I(\mathbf{Y}_1 \leqslant \mathbf{s}), I(\mathbf{Y}_{1+i} \leqslant \mathbf{t})) \\
& + \operatorname{cov}(I(\mathbf{Y}_1 \leqslant \mathbf{t}), I(\mathbf{Y}_{1+i} \leqslant \mathbf{s}))],
\end{aligned}
$$

and $\Gamma_* = \sup\{\Gamma(\mathbf{t}, \mathbf{t}); \mathbf{t} \in J\} > 0$, then

$$n^{-1} \log P(KS_n \geqslant \varepsilon_n) = -(n\varepsilon^2/2\Gamma_*)(1 + O_\epsilon + O_{n,\epsilon}).$$

Proof. For a set S, let $Ca(S)$ denote its cardinality. Given $\varepsilon > 0$, we define

$$B_\epsilon = \{0, \varepsilon^2, 2\varepsilon^2,..., [\varepsilon^{-2}]\,\varepsilon^2, 1\},$$
$$B_\epsilon^* = \{\mathbf{x}: \mathbf{x} \in B_\epsilon \text{ and } \Gamma(\mathbf{x}, \mathbf{x}) > (3c_{10})^{-2}\,\Gamma_*\}$$
$$B_\epsilon^{**} = B_\epsilon - B_\epsilon^*$$

and

$$\Gamma_{\epsilon^*} = \max\{\Gamma(\mathbf{t}, \mathbf{t}); \mathbf{t} \in B_\epsilon^*\},$$

where c_{10} is the constant appearing in Lemma 3.1, corresponding to $\|X_1\|_\infty = 1$. Thus $B_\epsilon = B_\epsilon^* \cup B_\epsilon^{**}$. Because of the uniform continuity of the function $\Gamma(\mathbf{x}, \mathbf{x})$ in $\mathbf{x} \in J$ and the condition $\Gamma_* > 0$, the set B_ϵ^* is nonempty for ε sufficiently small and $\Gamma_{\epsilon^*} = \Gamma_* + O_\epsilon$.

Given $\mathbf{x} \in J$, there exists an element $\mathbf{t} = (t_1, t_2,..., t_q)$ of B_ϵ which belongs to the set

$$\prod_{j=1,q} [t_j, (t_j + \varepsilon^2) \wedge 1].$$

This, together with the assumption that marginals are $U[0, 1]$, makes sure that

$$|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{t})| \leqslant q\varepsilon^2 \qquad \text{and} \qquad |\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{t}')| \leqslant q\varepsilon^2,$$

where $\mathbf{t}' = ((t_1 + \varepsilon^2) \wedge 1,..., (t_q + \varepsilon^2) \wedge 1)$. As a consequence of this and the monotone property of $\mathbf{F}_n$,

$$\Big|\sup_{\mathbf{t}\in J} |\mathbf{F}_n(\mathbf{t}) - \mathbf{F}(\mathbf{t})| - \max_{\mathbf{t}\in B_\epsilon} |\mathbf{F}_n(\mathbf{t}) - \mathbf{F}(\mathbf{t})|\Big| \leqslant q\varepsilon^2;$$

therefore

$$\begin{aligned} P(\max_{\mathbf{t}\in B_\epsilon^*} |\mathbf{F}_n(\mathbf{t}) - \mathbf{F}(\mathbf{t})| \geqslant \varepsilon_n) \leqslant P(KS_n \geqslant \varepsilon_n) \\ \leqslant P(\max_{\mathbf{t}\in B_\epsilon^*} |\mathbf{F}_n(\mathbf{t}) - \mathbf{F}(\mathbf{t})| \geqslant \varepsilon_n - q\varepsilon^2) \\ + P(\max_{\mathbf{t}\in B_\epsilon^{**}} |\mathbf{F}_n(\mathbf{t}) - \mathbf{F}(\mathbf{t})| \geqslant \varepsilon_n - q\varepsilon^2). \end{aligned} \tag{4.1}$$

We now estimate the probabilities appearing in (4.1) to arrive at the desired theorem.

By an appeal to Theorem 3,

$$\begin{aligned}
P(\max_{\mathbf{t}\in B_\epsilon^*} |\mathbf{F}_n(\mathbf{t})-\mathbf{F}(\mathbf{t})| \geqslant \varepsilon_n - q\varepsilon^2)
&\leqslant Ca(B_\epsilon^*)\max\{P(|\mathbf{F}_n(\mathbf{t})-\mathbf{F}(\mathbf{t})| \geqslant \varepsilon_n - q\varepsilon^2);\, t\in B_\epsilon^*\}\\
&\leqslant Ca(B_\epsilon^*)\exp(-n\varepsilon^2(1+O_\epsilon+O_{n,\epsilon})/2\Gamma_{\epsilon^*})\\
&\leqslant Ca(B_\epsilon^*)\exp(-n\varepsilon^2(1+O_\epsilon+O_{n,\epsilon})/2\Gamma_*). \qquad (4.2)
\end{aligned}$$

If $\mathbf{t}^o$ denotes a member of B_ϵ^* which maximizes $\Gamma(\cdot,\cdot)$ over B_ϵ^*, then

$$\begin{aligned}
P(\max_{\mathbf{t}\in B_\epsilon^*} |\mathbf{F}_n(\mathbf{t})-\mathbf{F}(\mathbf{t})| \geqslant \varepsilon_n) &\geqslant P(|\mathbf{F}_n(\mathbf{t}^o)-\mathbf{F}(\mathbf{t}^o)| \geqslant \varepsilon_n)\\
&= \exp(-n\varepsilon^2(1+O_\epsilon+O_{n,\epsilon})/2\Gamma(\mathbf{t}^o,\mathbf{t}^o))\\
&= \exp(-n\varepsilon^2(1+O_\epsilon+O_{n,\epsilon})/2\Gamma_*). \qquad (4.3)
\end{aligned}$$

Finally, the rightmost term of (4.1) is estimated using Lemma 3.1. Lemma 3.1 with $D=n(\varepsilon_n-q\varepsilon^2)/2c_{10}$, $Z=n\varepsilon^2/\Gamma_*$, and $B=(3c_{10})^{-2}\Gamma_*$ implies

$$\begin{aligned}
P(\max_{\mathbf{t}\in B_\epsilon^{**}} |\mathbf{F}_n(\mathbf{t})-\mathbf{F}(\mathbf{t})| \geqslant \varepsilon_n - q\varepsilon^2)
&\leqslant Ca(B_\epsilon^{**})\max\{P(|\mathbf{F}_n(\mathbf{t})-\mathbf{F}(\mathbf{t})| \geqslant \varepsilon_n - q\varepsilon^2);\, \mathbf{t}\in B_\epsilon^{**}\}\\
&= \exp(-n\varepsilon^2(1+O_\epsilon+O_{n,\epsilon})/\Gamma_*) \qquad (4.4)
\end{aligned}$$

for all ε small. The theorem now follows from (4.1)–(4.4).

To get a similar LDP bound for AD_n, we need to develop some notation first. For a function g defined on J, let $\|g\|_2 = (\int_J g^2(\mathbf{x})\,d\mathbf{x}))^{1/2}$. If l is a finite dimensional vector, $\|l\|$ denotes its Euclidean norm. Also, we define

$$[\Gamma\cdot h]_* = \sup\left\{\left\|\int_J \Gamma(\mathbf{s},\mathbf{t})\,h(\mathbf{s})\,h(\mathbf{t})\,g(\mathbf{s})\,d\mathbf{s}\right\|_2;\ g \text{ is a continuous function on } J \text{ with } \|g\|_2=1\right\},$$

$$\{\Gamma\cdot h\}_\epsilon = \{\Gamma(\mathbf{s},\mathbf{t})\,h(\mathbf{s})\,h(\mathbf{t})\}_{\mathbf{s},\mathbf{t}\in B_\epsilon},$$

a matrix of order $\mathrm{Car}(B_\epsilon)\times\mathrm{Car}(B_\epsilon)$,

$$[\Gamma\cdot h]_{\epsilon^*} = (\mathrm{Car}(B_\epsilon))^{-1}\sup\{\|\{\Gamma\cdot h\}_\epsilon\cdot l\|,$$

where $\mathbf{l}$ is a vector of dimension $\mathrm{Car}(B_\epsilon)$ with $\|\mathbf{l}\|=1\}$.

It is verified through the standard approximation of integrals with finite sum that $[\Gamma\cdot h]_{\epsilon^*} = [\Gamma\cdot h]_* + O_\epsilon$.

THEOREM 7. *If* $[\Gamma \cdot h]_* > 0$, *then*

$$n^{-1} \log P(AD_n \geqslant \varepsilon_n) = -(n\varepsilon^2/2[\Gamma \cdot h]_*)(1 + O_\epsilon + O_{n,\epsilon}).$$

Proof. Approximation of AD_n^2 by an average over the set B_ϵ implies

$$\begin{aligned} &\left| AD_n^2 - (\mathrm{Car}(B_\epsilon))^{-1} \sum_{\mathbf{t} \in B_\epsilon} (\mathbf{F}_n(\mathbf{t}) - \mathbf{F}(\mathbf{t}))^2 h^2(\mathbf{t}) \right| \\ &\quad \leqslant \sup\{|(\mathbf{F}_n(\mathbf{s}) - \mathbf{F}(\mathbf{s}))^2 h^2(\mathbf{s}) - (\mathbf{F}_n(\mathbf{t}) - \mathbf{F}(\mathbf{t}))^2 h^2(\mathbf{t})|; \\ &\qquad |\mathbf{s} - \mathbf{t}| \leqslant \varepsilon^2, \mathbf{s}, \mathbf{t} \in J\} \\ &\quad \leqslant c_{12} \sup\{|\mathbf{F}_n(\mathbf{s}) - \mathbf{F}(\mathbf{s}) - \mathbf{F}_n(\mathbf{t}) + \mathbf{F}(\mathbf{t})| + |h^2(\mathbf{s}) - h^2(\mathbf{t})|; \\ &\qquad |\mathbf{s} - \mathbf{t}| \leqslant \varepsilon^2, \mathbf{s}, \mathbf{t} \in J\}, \end{aligned}$$

where $|\mathbf{s} - \mathbf{t}| \leqslant \varepsilon^2$ means that the differences between corresponding coordinates are $\leqslant \varepsilon^2$. An application of Corollary 3.1 shows

$$\begin{aligned} &n^{-1} \log P\left((\mathrm{Car}(B_\epsilon))^{-1} \sum_{\mathbf{t} \in B} (\mathbf{F}_n(\mathbf{t}) - \mathbf{F}(\mathbf{t}))^2 \mathbf{h}^2(\mathbf{t}) \geqslant \varepsilon_n^2 \right) \\ &\quad = -(n^2\varepsilon^2/2[\Gamma \cdot h]_{\epsilon^*})(1 + O_\epsilon + O_{n,\epsilon}) \\ &\quad = -(n\varepsilon^2/2[\Gamma \cdot h]_*)(1 + O_\epsilon + O_{n,\epsilon}). \end{aligned}$$

Because h is a continuous function on the compact set J,

$$\sup\{|h^2_\epsilon(\mathbf{s}) - h^2(\mathbf{t})|; \mathbf{s}, \mathbf{t} \in J, |\mathbf{s} - \mathbf{t}| \leqslant \varepsilon^2\} = O_\epsilon.$$

Hence it only remains to be seen that the remainder

$$c_{13} \sup\{|\mathbf{F}_n(\mathbf{s}) - \mathbf{F}(\mathbf{s}) - \mathbf{F}_n(\mathbf{t}) + \mathbf{F}(\mathbf{t})|; \mathbf{s}, \mathbf{t} \in J, |\mathbf{s} - \mathbf{t}| \leqslant \varepsilon^2\} \tag{4.5}$$

is $(*)$. This is done using the Bonferroni inequality and Lemma 3.1. If $|\mathbf{s} - \mathbf{t}| \leqslant \varepsilon^2$,

$$\begin{aligned} &V(I(\mathbf{Y}_1 \leqslant \mathbf{s}) - I(\mathbf{Y}_1 \leqslant \mathbf{t})) + 2 \sum_{i=1,m-1} \mathrm{cov}(I(\mathbf{Y}_1 \leqslant \mathbf{s}) - I(\mathbf{Y}_1 \leqslant \mathbf{t})), \\ &I(\mathbf{Y}_{1+i} \leqslant \mathbf{s}) - I(\mathbf{Y}_{1+i} \leqslant \mathbf{t}) \leqslant 2mq\varepsilon^2. \end{aligned} \tag{4.6}$$

The dependence of (4.6) on ε^2 is exploited through Lemma 3.1 to prove that (4.5) is $(*)$.

ACKNOWLEDGMENT

Referee's suggestions helped improve the presentation of this paper.

REFERENCES

[1] BABU, G. J. AND SINGH, K. (1978). On deviation between empirical and quantile processes for mixing random variables. *J. Multivar. Anal.* **8**, 522–549.

[2] BAHADUR, R. R. (1960). On the asymptotic efficiency of tests and estimates. *Sankhyā* **22**, 229–252.

[3] BAHADUR, R. R. (1969). *Some Limit Theorems in Statistics*. SIAM Publications, Philadelphia, Pa.

[4] CHANDA, K. C. (1972). On estimation of tail end probabilities of the sample mean of linear stochastic processes. *Ann. Math. Statist.* **43**, 1680–1685.

[5] IBRAGIMOV, I. A. (1962). Some limit theorems for stationary processes. *Theory Probab. Appl.* **7**, 349–382.

[6] RUBIN, H. AND SETHURAMAN, J. (1965). Probabilities of moderate deviations. *Sankhyā Ser. A* **27**, 325–346.